Further Development and Applications of GeoChip 3.0 for Microbial Community Analysis

Zhili He[1,6], Ye Deng[1,6], Meiying Xu[1,6], Liyou Wu[1,6], Yuting Liang[1,6], Joy D. Van Nostrand[1,6], Christopher Hemme[1,6], Terry J. Gentry[2], **Jost Liebich**[3], Adam P. Arkin[4,6], Terry C. Hazen[5,6], and Jizhong Zhou[1,6]

[1]The University of Oklahoma, Norman, OK, [2]Texas A&M University, College Station, TX, [3]Agrosphere Institute (ICG-4), Forschungszentrum Jülich GmbH, Jülich, Germany, [4]Physical BioScience Division, Lawrence Berkeley National Laboratory, Berkeley, CA, [5]Earth Science Division, Lawrence Berkeley National Laboratory, Berkeley, CA, [6]Virtual Institute for Microbial Stress and Survival, http://vimss.lbl.gov

**ESPP2** — Environmental Stress Pathway Project — Q-160

http://vimss.lbl.gov/

Abstract

Microarrays constructed with the genes encoding key enzymes involved in various biological and geochemical processes are referred to as functional gene arrays (FGAs), or GeoChip. On the basis of GeoChip 2.0 (He et al. 2007), which contains 24,243 oligonucleotide (50mer) probes and covers > 10,000 gene sequences in >150 functional groups involved in nitrogen, carbon, sulfur and phosphorus cycling, metal reduction and resistance, and organic contaminant degradation, a new generation of GeoChip (GeoChip 3.0) has been developed. GeoChip 3.0 has several new features compared to GeoChip 2.0. First, GeoChip 3.0 cover about 47,000 gene sequences for 292 gene families, and such a coverage allows us to obtain more information about microbial communities and analyze more diverse environmental samples. Second, the homology of automatically retrieved sequences by key words is verified by HUMMER using seed sequences so that any potential unrelated sequences are removed. Third, GeoChip 3.0 includes phylogenetic markers, such as *gyrB*. Fourth, a software package (including databases) has been developed for sequence retrieval, probe and array design, probe verification, array construction, array data analysis, information storage, and automatic update, which greatly facilitate the management of such a complicated array, especially for future update. Fifth, universal standards have been implemented in GeoChip 3.0 so that data normalization and comparison of different samples can be conducted. Finally, genomic standards are also used to quantitatively analyze gene abundance. GeoChip was used to analyze effects of elevated CO_2 on belowground microbial communities at BioCON. The results demonstrated that GeoChip 3.0 could provide more capability for studying biogeochemical processes and functional activities of microbial communities important to human health, agriculture, energy, global climate change, ecosystem management, and environmental cleanup and restoration. It is also particularly useful for providing direct linkages of microbial genes/populations to ecosystem processes and functions.

GeoChip 3.0 Development

Work flow for GeoChip 3.0 design, construction and data analysis
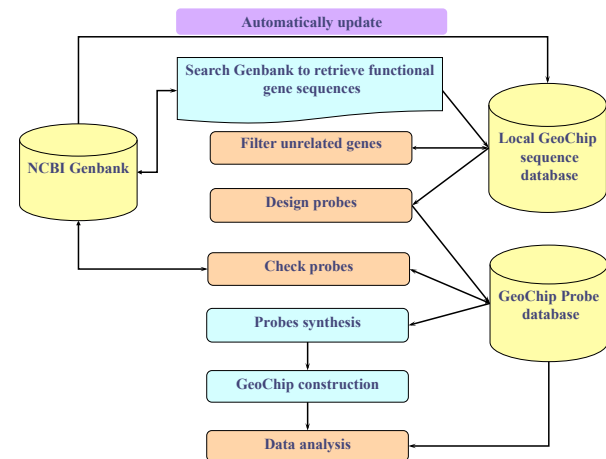
ACKNOWLEDGEMENT

New features of GeoChip 3.0

- GeoChip 3.0 is more comprehensive, and it contains >24,500 probes and covers about 47,000 (~10,000 for GeoChip 2.0) gene sequences of 292 gene families (~150 gene family on GeoChip2.0). Thus, GeoChip 3.0 will be more representative.

- The homology of automatically retrieved sequences by key words is verified by HUMMER using seed sequences so that unrelated sequences ca be removed.

- A software package (including databases) has been developed for sequence retrieval, probe and array design, probe verification, array construction, array data analysis, information storage, and automatic update, which greatly facilitate the management of such a complicated array, especially for future updates.

- GeoChip has implemented a universal standard, which can compare different samples, and normalize data.

- GeoChip 3.0 implements a genomic control/standard, which can quantitatively analyze functional gene data.

- Automatic update greatly facilitates the management of such a complicated functional gene array.

### The summary of GeoChip 3.0 probe and sequence information

Gene category	No. of gene categories	No. of sequences for probe design	Total no. of probes designed	Total no. of CDS covered
Carbon degradation	31	9839	2720	4737
Carbon fixation	5	3378	898	1806
Methane reduction and oxidation	3	4182	254	434
Metal resistance and reduction	41	16825	4917	10458
Nitrogen cycling	13	27162	3561	6892
Organic remediation	190	31236	8815	16948
Phosphorus utilization	3	1441	599	1212
Sulfur cycling	3	4296	1328	1773
Energy process	2	901	413	449
Others (e.g. *gyrB*)	1	7957	1164	2251
Total	**292**	**107217**	**24669**	**46960**

Universal standards implemented in GeoChip 3.0

- Gene probes
- Gene probes + the reference probe
- Reference probe

- Each functional gene probe is co-spotted with small amount of the reference probe.
- Universal standards allow to normalize and compare different microbial communities.

Quantitative standards implemented in GeoChip 3.0

Target gene probes / Control gene probes

- Seven hyperthermophiles were selected, and each has been designed 96 gene-specific probes.
- Different amounts (0.02-250 pg/gene) of each gDNA were spiked into every sample.
- Control and gene-specific probes are spotted separately.
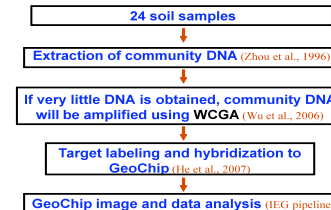- Genomic standards enable us to quantitatively measure functional gene abundance or activity.
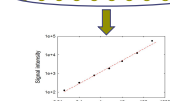
Applications of GeoChip 3.0 to analyze effects of elevated CO_2 on belowground microbial communities at BioCON

BioCON (biodiversity, CO2, and nitrogen) site and sampling
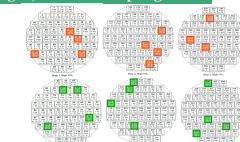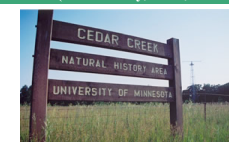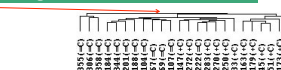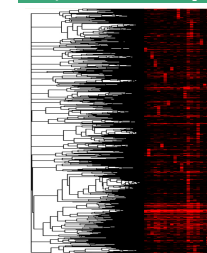

24 soil samples were taken from 12 plots of high CO_2 (560 µmol/mol, red) rings (1, 3 and 5) and 12 plots of ambient CO_2 (368 µmol/mol, green) rings (2, 4 and 6) in July 2007 at the BioCON site as shown above.

Major steps for microbial community analysis using GeoChip 3.0

24 soil samples

↓

Extraction of community DNA (Zhou et al., 1996)

↓

If very little DNA is obtained, community DNA will be amplified using WCGA (Wu et al., 2006)

↓

Target labeling and hybridization to GeoChip (He et al., 2007)

↓

GeoChip image and data analysis (IEG pipeline)

Analysis of effects of CO_2 on belowground microbial communities

Number of the functional gene detected

Sample	Ambient CO_2	Elevated CO_2
Total gene detected	2541	2850
C-cycle	289	324
Energy procession	43	47
CH4 cycle	20	23
Metal resistance	497	587
N-cycle	283	321
Organic remediation	1151	1266
Phosphorus	56	65
Sulphurs	118	128

CONCLUSIONS

1. GeoChip 3.0 has been constructed with more than 24,500 probes and covers about 47,000 gene sequences in 292 gene families. To our knowledge, this is the most comprehensive functional gene array currently available for environmental studies.

2. GeoChip 3.0 has many new features from rapid sequence retrieval and verification, probe design and validation, data comparison and normalization, quantitative analysis of functional gene abundance, to future updates.

3. GeoChip 3.0 has been used to analyze effects of elevated CO_2 on belowground microbial communities, and the results demonstrate that GeoChip is a powerful tool for analysis of geochemical, biological and environmental processes and associated microbial communities.